

The International Journal of Digital Curation

Issue 1, Volume 2 | 2007

Digital Curation Centre – Phase Two

Helen Hockx-Yu,
PLANETS Project Manager,
The British Library

June 2007

Summary

An overview of the Digital Curation Centre since its establishment in 2004, including accomplishments of Phase One, aims and objectives of Phase Two, and how DCC relates to JISC strategy.

Introduction

The ease with which digital information can be created, combined with the huge increase in computer power and network bandwidth, has led to the proliferation of a vast amount of ‘born-digital’ data, especially in science and engineering, where petabytes of data are being generated by scientific instruments on a daily basis. In parallel with this there is an exponential growth of diverse and complex electronic publications and resources, which increasingly form the primary corpus of record for research and education.

To address this data deluge and the potential loss of valuable research outputs, the Digital Curation Centre (DCC) was established in 2004 to help ensure that data generated today can survive changes in technology and remain accessible to users in the future. As a joint venture between the Joint Information Systems Committee (JISC)¹ and the Research Councils e-Science Core Programme², the DCC was tasked to lead research and development into key areas of digital curation for data and publications, and to pilot the development of generic support services for curating digital data and research results for current and future users.

DCC Phase One

Phase one of the DCC involved funding of over £3 million from the JISC which was completed in February 2007. Over the three years of Phase One, the DCC has created a pool of resources for the community and produced an impressive slate of outputs, including conferences, workshops, seminars, briefing papers, interviews, case studies, technology and standards watch papers, a catalogue of curation tools, a peer-reviewed international journal, and a *Curation Manual*. The DCC has in addition played a leading role in work related to the audit and certification of digital repositories and released a pilot registry / repository of Representation Information³ which forms a key component of any preservation infrastructure.

The DCC’s Research Programme, supported by over £1 million from the e-Science Core Programme till February 2008⁴, investigates issues identified as crucial to digital curation, such as database annotation and provenance, data transformation, integration and publishing, socio-economic and legal issues, metadata extraction and curation. This has already resulted in many high-quality research papers, which have been accepted by leading peer-reviewed journals. An externally moderated evaluation, carried out between February and August 2006, concluded that: the DCC is necessary to the well-being of UK research, enterprise and culture; has established itself with high-quality outputs; and should be encouraged to expand its activities *selectively*. (Brophy & Fry, 2006). The DCC has also made a strong international presence, which was illustrated by the fact that it was named one of ten most promising digital preservation initiatives by the US Research Libraries Group’s DigiNews (RLG DigiNews, 2005).

¹ Joint Information Systems Committee (JISC): <http://www.jisc.ac.uk>

² UK e-Science Programme: <http://www.rcuk.ac.uk/escience/default.htm>

³ Defined by Open Archival Information System (OAIS) Reference Model to refer to all information required to access and interpret a digital object.

⁴ It has not been expected that the funding for the Research Programme will be renewed in the same way. However, prospective bids and other funding sources will enable the research agenda to be taken forward.

Planned Improvements

While confirming DCC's achievement, the evaluation also revealed a number of areas in which improvement is needed so that the DCC can become more effective and have greater impact. These include more active engagement with the critical community, the need to develop products and services for different parts of the community and better integration with other curation services and data centres.

The outcome of the evaluation was an important reference for the JISC Support of Research Committee (JSR), which oversees JISC's funding for activities that support the needs of researchers. Making available £3.4 million, the JSR decided in September 2006 to fund Phase Two of the DCC for another three years. The decision-making process did not involve lengthy consideration as to whether or not to fund the DCC again as its continued existence was clearly in line with the JISC Strategy. The focus was more on the shape and scope of Phase Two, in particular how the DCC can engage with researchers more effectively.

The Committee strongly requested that DCC Phase Two have a specific focus on research data, and that the areas to be addressed should be those which were identified as in need of improvement by the evaluation exercise. In addition, the Committee would like to see DCC involved in staff development for digital curation skills, both for internal staff and researchers in the community. Sustainability is also high up on the agenda; a sustainability model needs to be developed. In addition, the future shape and scope of the DCC will need to be explored further.

The Context of Phase Two

The growing importance of digital research data has been recognised by funding bodies around the world and there is a global trend across funding bodies to require that research outputs are not only made publicly available but also managed properly in order to maintain long-term access. Comparing the wider context in which the DCC was established three years ago, there is far greater awareness of the need for digital curation and a much clearer understanding of the relevant issues associated with it. Many initiatives have been undertaken to investigate policies and strategies at national level.

The UK Office of Science and Innovation (OSI) established a high-level e-infrastructure steering group which recognised preservation and curation as key components in the future national e-infrastructure for research. A sub-working group was consequently formed to map out relevant developments, gaps and challenges and to assess funding requirements. The working group reported in early 2007 and recommended over £60 million funding for a Digital Preservation and Curation Research and Development Programme and a Persistent National Information Infrastructure Development Programme (Beagrie, [2007](#)).

In the US, the National Science Board for example formed the Long-lived Data Collections Task Force to address the policy issues related to supporting the collection and curation of research data (National Science Foundation, [2005](#)). Data are an important component of the cyberinfrastructure envisaged by the US National Science Foundation for 21st Century Discovery, whose vision is for science and engineering digital data to be: routinely deposited in well-documented form; regularly and easily

consulted and analysed by specialist and non-specialist alike; and openly accessible while being suitably protected and reliably preserved (National Science Foundation Cyberinfrastructure Council, 2007). The Australian government has also established a similar strategy for a collaborative research infrastructure.⁵

In addition to policy developments, significant investment has been made in digital preservation and curation research. DCC Phase Two runs in parallel with JISC's Digital Repositories and Preservation Capital Programme⁶, which has made £14 million available to develop preservation infrastructure and capacity for digital repositories. The European Commission funded three large-scale digital preservation projects in 2006 under Call 5 of its Information Science and Technologies Framework Programme 6, involving funding of over 17 million Euros in total.⁷

DCC Phase Two

DCC Phase Two is not only a response to the external environment, it is also seen as mechanism to deliver the JISC strategy and support the research community. JISC believes that there are considerable benefits in providing leadership, advice and guidance to the sector through advisory services such as the DCC, which are also a valuable source of input to JISC on the evolving needs of the community. DCC is formally a project under JISC's development programme but has been listed as an example of advisory services in the JISC Strategy 2007-2009.⁸ This illustrates the strategic importance of the DCC but also sets a great challenge to sustain the service beyond the current funding commitment.

The vision of DCC Phase Two is for it to become the centre of excellence in digital curation and preservation in the UK, the authoritative source of advocacy and expert advice and guidance to the community, the key facilitator of an informed research community with established collaborative networks of digital curators and the service provider of a wide range of resources, software, tools and support services. A set of objectives has been drawn up to achieve this vision:

- Provide strategic leadership in digital curation and preservation for the UK research community, with particular emphasis on science data
- Influence and inform national and international policy
- Provide advocacy and expert advice and guidance to practitioners and funding bodies
- Create, manage and develop a highly effective suite of resources and tools
- Raise the level of awareness and expertise amongst data creators and curators, and amongst other practitioners with a curation role
- Strengthen community curation networks and collaborative partnerships
- Continue the DCC's strong association with the Research Programme⁹

⁵ National Collaborative Research Infrastructure Strategy (NCRIS): <http://www.ncris.dest.gov.au/>

⁶ JISC Repositories and Preservation Capital Programme:
http://www.jisc.ac.uk/whatwedo/programmes/programme_rep_pres.

⁷ The three projects are: PLANETS (<http://www.planets-project.eu/>), CASPAR (<http://www.casparpreserves.eu/>) and DigitalPreservationEurope (<http://www.digitalpreservationeurope.eu/>).

⁸ "JISC Strategy 2007-2009", available at http://www.jisc.ac.uk/media/documents/about_us/strategy/jisc_strategy_2007-2009.pdf

⁹ Digital Curation Centre (2007) "About the DCC", available at <http://www.dcc.ac.uk/about/>

The key to the DCC's success is that it remains relevant to the community. The way the DCC defines its community and its understanding of that community's needs are fundamental to DCC's service delivery. Data creators, data curators and data re-users were defined as the main audiences in DCC Phase One. A lesson learnt was that these segmentations are too broad-brushed and did not take into account the different levels of awareness, experience and expertise related to digital curation that exist within each of the aforementioned audiences. In Phase Two, the DCC proposed to refine the definition of its target community, to include those who are more experienced, emerging and those who are still unengaged, and to develop different communications and services to benefit these groups. The DCC has also responded by re-organising its internal management structure and setting up a new Community Development team to carry out community building activities and to focus on community links at a higher expertise level. Furthermore, plans have been made to recruit staff with a higher level of curation expertise.

The DCC has not been funded to handle data directly or operate any data repositories. It is therefore crucial that the DCC works with existing data centres and curation services to benefit from their practical expertise. This is an identified area for improvement in Phase Two. Activities will be organised around the Disciplinary Approaches to Sharing, Curation, Re-use and Preservation (SCARP) Project¹⁰, funded under the JISC e-Research Capital Programme to address disciplinary data management, sharing and curation practices, and to enhance interaction and exchange between existing data centres. In addition, links will be sought with Research Council-funded eScience programmes and projects in order to embed good data management and curation practices.

The DCC's UK-wide remit has given it a unique position to facilitate a coherent approach to digital curation and preservation for the research community. A goal in Phase 2 is to engage with the Research Councils and other key players and produce a Digital Curation Roadmap for the UK to outline a long-term strategy. This should enable us to achieve a shared view of the roles and responsibilities for digital curation at national and institutional level and would be a great contribution to the UK research community.

Conclusion

The DCC has a large number of challenges ahead of it. Building on the ground work and achievement in Phase One while supported by a group of professional and enthusiastic staff, the DCC should look forward to further success in its support of the research community.

References

- Brophy, P., & Fry, J. (2006). *Externally moderated reflective self-evaluation: Report*. Edinburgh, UK: Digital Curation Centre. Retrieved July 11, 2007 from http://www.dcc.ac.uk/docs/DCC_Evaluation_Report_Final.pdf
- Beagrie, N. (2007). *e-Infrastructure strategy for research: Final report from the OSI Preservation and Curation Working Group*. London: Office of Science and Innovation (OSI). Retrieved July 11, 2007 from <http://www.nesc.ac.uk/documents/OSI/preservation.pdf>

¹⁰ Disciplinary Approaches to Sharing, Curation, Re-use and Preservation (SCARP) <http://www.dcc.ac.uk/scarp/>

National Science Foundation. (2005). *Long-lived digital data collections enabling research and education in the 21st century*. National Science Board report No. NSB-05-40. Retrieved July 11, 2007 from:

<http://www.nsf.gov/pubs/2005/nsb0540/>

National Science Foundation Cyberinfrastructure Council. (2007). Cyberinfrastructure vision for 21st century discovery. Retrieved July 11, 2007 from:

<http://www.nsf.gov/pubs/2007/nsf0728/nsf0728.pdf>

RLG DigiNews. (2005). Watch this space: ten promising digital preservation initiatives. *RLG DigiNews*, Vol 9, No. 4. Retrieved July 11, 2007 from

http://www.rlg.org/en/page.php?Page_ID=20744